Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.6, No.36, 2015

Do Scaffolded Supports between Aspects of Problem Solving Enhance Assessment Usability?

Jan D. McCoy[1*] Jenelle Braun-Monegan[2] Leanne Bettesworth[3] Gerald Tindal[4]

1. Oregon Department of Education, 255 Capitol Street NE, Salem, Oregon 97310 United States
2. Pacific University, School of Learning and Teaching, 40 E Broadway # 250, Eugene, OR 97401 United States
3. University of Calgary, Werklund School of Education, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada
4. University of Oregon, College of Education, 1215 University of Oregon, Eugene, OR 97403-1215 United States

Abstract

While problem solving as an instructional technique is widely advocated, educators are often challenged in effectively assessing student skill in this area. Students failing to solve a problem might fail in any of several aspects of the effort. The purpose of this research was to validate a scaffolded technique for assessing problem solving in science and social studies at the middle school level. This technique attempts to isolate three aspects of problem solving (data collection, analysis and display, and interpretation) and to measure each aspect separately. Problem solving measures were developed in both science and social studies. These were administered both fall and spring to determine student skill in problem solving and to measure growth in problem solving skill over time and differential skill across grades (6 through 8). Segmented tasks were scaffolded between segments to circumvent the interdependency of elements of the problem solving process. It was determined the measures were successful in supporting students who had difficulty across segments within a single problem solving task and student problem solving skills could be evaluated effectively using the results of the measure.

Keywords: scaffolding, assessment, problem solving

1. Introduction

As we move into the twenty-first century, problem solving as a skill is gaining in emphasis and prominence among educators (Cho, Caleon, & Kapur, 2015). Unfortunately these educators face the challenge of determining how to assess students' problem solving skill levels. While assessments often approach problem solving as a unitary activity, successful assessments must acknowledge the phased approach demanded by problem solving (Jonassen, 2014). This article will describe the steps undertaken to design and develop four measurement instruments (two in science and two in social studies) intended to measure growth in student problem-solving skill. We will then explain efforts to validate the instruments for teacher decision-making in evaluating students' problem-solving skills. The instruments were intended to measure: (a) growth fall to spring for a single student progressing through a school year, (b) differentiation among student performances within a single grade, and (c) differential achievement across grades 6 through 8. Measuring problem solving has always been complicated as it requires many interdependent skills. Trying to identify where in the problem solving process students are having difficulty has not yet been accomplished well. These measures developed were unique in their approach to problem solving in that they attempted to accommodate students with shortcomings in within their understanding of the problem solving steps by providing scaffolding for those students in addressing the problems provided so they could continue to progress through the problem.

It is widely accepted that problem solving (also known as problem-based learning) is a useful technique for student instruction (Jonassen, 2011; Segers, Van den Bossche, & Teunissen, 2003). Under this instructional model, students are presented with ill-structured problems and challenged to arrive at a reasonable decision among proposed outcomes or a solution to the problem. The difficulty with this approach has been that, while it is easy enough to use this model in instruction, it is more difficult to assess student proficiency in solving such problems. This is true for two reasons, (a) most instruction undertaken in this model is done in groups rather than individually and (b) ill-structured problems present such variability in response and are so dependent on correct decisions at several points in the process that points of student failure in problem solving are difficult to isolate

(Dochy, Segers, Van den Bossche, & Gijbels, 2003). Problem solving is a complex and interdependent task (Jonassen, 2011). It was proposed that the scaffolding could assist students moving through the task. Testing this proposition demands clear definitions of both *problem solving* and *scaffolding.*

Problem solving is a process incorporating both knowledge and skills. This includes identifying a problem, defining the source of the problem, collecting information, and exploring solutions to the problem. Exploring potential solutions requires the ability to make decisions using information found in the investigation (Brophy, 1998). The steps of problem solving listed above are essentially repeated in the National Science Standards as proposed by the National Research Council (1996, 2000). Problem solving steps in science and social studies are variously described in the literature but can be summarized as a four-step process: (a) definition of the problem and design of a data collection plan and desired dataset, (b) data collection, (c) data analysis and display, and (d) data interpretation.

The concept of scaffolding, based on the construction technique where a temporary framework is erected to help workers reach areas of the project that would otherwise be out of their reach (Flick, 2003), is similarly applied in education. That is, the scaffolding included in the measure assists students who would have difficulty reaching a solution to a proposed problem. The word scaffolding was first used in an educational sense by Wood, Bruner, and Ross (1979) who described it as, "…a kind of 'scaffolding' process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts" (p. 90). Scaffolding has since been used as an instructional technique to help students manage complex problems in subject areas such as reading, mathematics, science, and problem solving, and is now thought of as an effective teaching strategy (De León, 2012; Joseph, 2002; Sandoval & Reiser, 2004; White & Frederiksen, 1998; Wolf, Brush, & Saye, 2003). However, scaffolding has not been as widely used in assessment as it has in instruction. Our proposed problem solving measure is unique because of its incorporation of scaffolding between each segment of the measure. The research described here attempts to apply this successful instructional technique to assessment of problem solving.

The validation of these measures of problem solving in science and social studies at the middle school level required that they be broad enough to discern growth among 6th, 7th, and 8th grade students as well as growth within a single student over the course of a school year. This technique involved not only a measure of structured problem solving but also the use of scaffolding to allow students who were weak in one area of problem solving to address other areas without hindrance. It was hypothesized that students facing a problem solving task without scaffolding would have difficulty with later segments of the task were they to fall short in early segments. This would be the likely result of interdependency across the task. The scaffolding provided was intended to support students who were unsuccessful in early segments of the task in their attempts to address later segments.

Our research question was, "Does providing students with support from one aspect of problem solving to another (scaffolding) enhance our assessment of student's problem solving skill?"

2. Research Methodology

This study is a quantitative exploration of an assessment technique to measure scientific inquiry. These instruments were designed to elicit and evaluate multiple inquiry skills from students using a phased gate processes. Typically scientific and social studies inquiry requires multiple interdependent skills, however, this structured phased gated assessment aimed at deconstructing the multiple interdependent skills into its subcomponents: (a) definition of the problem and design of a data collection plan and desired dataset, (b) data collection, (c) data analysis and display, and (d) data interpretation. To achieve this deconstruction once students completed a component they were then moved to the next component while being provided with both the information they generated during the previous component as well as the information that should have been generated in the component. This phase gated approach allowed students to face each segment or component as an independent measure and provided the opportunity to identify shortcomings and strengths of students' inquiry skills on all three tasks.

In unscaffolded assessments, those students who did not perform well on early tasks would not have the information necessary to succeed on subsequent tasks. This scaffolded assessment was designed to support students who may have had trouble with early tasks in the assessment by providing them with correct versions of early tasks (along with the work they produced) when addressing later tasks. This provided support to those who might not be competent with one aspect of problem solving so that they could still address other aspects of the assessment successfully. The second component of each instrument presented the data necessary to solve the

problem thus supporting the students who were unsuccessful with component one (identifying the data needed) to be able to address component two (data display and analysis). Similarly, the third segment of each instrument provided a complete data analysis (addressing the task from segment two) and asked students to respond with an interpretation of the data and support their decision. This scaffolded approach allowed students to face each segment as an independent measure and provided the opportunity to identify shortcomings on the part of students without eliminating or hindering those who had difficulty with segments one or two.

2.1 Participants

Convenience sampling was used where all science and social studies teachers at a local middle school were invited to participate. Sixteen teachers from a single middle school in a suburban district in the Pacific Northwest responded to the research team willing to participate in this project. Their 421 students in science and social studies classes (66 at sixth grade, 186 at seventh grade, and 169 at eighth grade) served as participants. These students were predominantly white (approximately 88%) as is typical of the region. Males and females were nearly evenly divided across grades. Approximately 40% of the population received free and reduced lunch while approximately 12% received special education services.

2.2 Data Collection Procedure

Researchers with both teaching and assessment experience developed four measures, two each in science and social studies, targeting problem solving skills (see Appendix A). The assessment framework broke the larger construct of problem solving down into its underlying components as illustrated in our logic model. Each component addressed a single aspect of problem-solving: (Component 1) design of a data collection plan and desired dataset, (Component 2) data analysis and display, and (Component 3) data interpretation. Students were not required to collect data (the second component outlined in our logic model) as data collection is a long complicated process that cannot be distilled into a forty five minute assessment session. Instead students were provided a data set for each of the problems after they designed a data collection plan.

In keeping with our assessment framework, the first component of the test presented students with a scenario that described a situation requiring a decision. This component of the test offered only enough information to outline the problem and to elicit a data collection plan and description of desired dataset. Students responded indicating what data would be needed to address the problem and how each data item needed would be useful in arriving at a solution.

In the second component, students were presented with a collection of data that included both relevant and irrelevant data sources. The data set fully described the situation under consideration. Students were asked to identify among the data presented those needed in solving the problem and those extraneous to the effort. Students were also tasked with organizing and displaying relevant data in a format supporting interpretation. Acceptable formats included charts, graphs, and/or tabular displays.

Component three offered students a collection of data displays including different tables and/or graphs as appropriate. Students interpreted these displays to determine which were relevant and how they contributed to developing a solution. Students were also tasked with providing a solution to the problem and justifying that solution using the data provided. With each segment, students had access to their previous day's work.

Topics were selected to be appropriate to the middle level curriculum but not included in instruction at the research site. This meant that students could reasonably be expected to understand and address the problem but that they did not receive instruction specific to the topic addressed. Both of the science problems dealt with the concept of *reproduction and growth in organisms* with the attributes of nutrition, environment, and population. Science form A asked students to determine the best medium for raising redworms for composting while form B asked students to determine the optimal soil type and watering pattern for raising peas. The social studies problems dealt with the concept of *place* with the attributes of geography, culture, and economics. Social studies form A asked students to choose between living in Eugene, Oregon (a midsized city) and Portland, Oregon (a more metropolitan area) with form B asking them to choose between two sites for the location of a gravel pit based on economic and ecological concerns (see Appendix A).

Each student faced two of the instruments, one in science (form A or form B) and the second in social studies (form A or form B) as a pretest in the fall and the alternate combination of measures (counterbalanced) as a posttest in the spring. For the fall administration, students were randomly assigned to a combination of

instruments by classroom. Each possible arrangement of the four measures, in any of four possible permutations, was administered to a randomly selected subset of students based on classroom assignment and stratified by grade (See Table 1). In each instance, instruments were administered over the course of 3, 30-minute sessions with each segment of the test administered on separate but contiguous days. To maintain fidelity of administration all segments were proctored by graduate students from the research team rather than by the classroom teachers (teachers remained present in the classroom).

Table 1. Order of Administration of Instruments

Classroom Assignment	Fall Administration		Spring Administration	
1	Science-A	Social studies-A	Science-B	Social studies-B
2	Science-A	Social studies-B	Science-B	Social studies-A
3	Science-B	Social studies-A	Science-A	Social studies-B
4	Science-B	Social studies-B	Science-A	Social studies-A

2.3 Scoring

Trained reviewers, blind to the students' grade level and other demographic data, scored student work. These reviewers used a scoring rubric unique to each segment of the measure. The scorers developed the rubric as an element of their training. Using 20 student responses from the fall administration as anchors, sample papers were ranked for quality and then examined for specific traits identifying and thoroughly describing each of six score points for each segment of the assessment. Student work was scored on a range from 0 for no attempt to 5 for exceptional work (see tables 2, 3, and 4). For the three segments combined, a student might receive a score ranging from 0 for no attempt on any of the three segments to 15 for exemplary performance on each segment. Student performance was analyzed by segment in an attempt to identify specific shortcomings in student problem solving. Validating this assessment technique depends on our finding subgroups of students who did not perform well on early segments but performed well on later segments of the test. Information on student's skills demonstrated during components 2 and 3 are what would typically be lost in a combined assessment.

Table 2.Dimensions for scoring student work for task 1

Score	Description of student performance
5	Essential information is included. Association of information to cause/effect. All attributes are included (nutrition, environment, population/geography, culture, economics). Includes questions or appropriate steps for further study.
4	Most essential information is included with cause and effect. Explicit mention of at least 1 attribute. Weak organization to attribute. Includes questions or appropriate steps for further study.
3	List of questions or steps, partially incomplete. Not well organized. Implicit attributes or explicit attributes with no cause and effect or cause and effect for 1 attribute.
2	Questions or recounting of information from task. Attributes not present or not connected to cause/effect or cause/effect missing.
1	Task not addressed. Minimal response.
0	Blank. No attempt at response.

Table 3.Dimensions for scoring student work for task 2

Score	Description of student performance
5	Visually displays tables or graphs. Clear independent and dependent variables. Accurate labels and titles included. Includes explanation for leaving out spurious data from tables and graphs.
4	Graphs or tables include spurious data. All pertinent data displayed. Accurate labels included may not include units or titles. Data appear to be accurately displayed. Acknowledges missing data.
3	Graphs or tables created. Data included are adequate to draw a conclusion relative to task. Labels not complete on graphs and tables.
2	Few variables are presented as graph or table data. Displays are inadequate for a conclusion. Variables misidentified or not identified.
1	No visual display of data (narrative may be included) OR Task not addressed.
0	Blank. No attempt at response.

Table 4. *Dimensions for scoring student work for task 3*

Score	Description of student performance
5	Obvious comparison and contrasts. Lots of specific and accurate examples from tables and graphs. Clear statement supporting the decision. Clear rationale for selection of criteria. Acknowledges elimination of data from decision-making regardless of explanation.
4	Decision clearly based on data. Some specific examples from tables and graphs.
3	Clear reference to data tables and charts. Decision clearly based on data. Some attempt at comparison and contrast.
2	Clear decision not supported by data. May quote data as supporting information.
1	No decision. May quote data as supporting information. Some response but not addressing the task.
0	Blank. No attempt at response.

2.4 Validating the gated phrase technique

As an initial step in validating these instruments, it was necessary to confirm that the segmenting offered by each of the components provided a measure of support to at least some of the students, thereby justifying the usefulness of the technique. Our premise in this effort held that, without the scaffolding, students who did not perform well on either segment one or segment two of the measure could not be expected to perform well on subsequent tasks. For example, students who did not correctly define the needed dataset could not hope to adequately analyze and display the data for subsequent interpretation without the provided scaffolding. Validating this scaffolding depended on our finding subgroups of students who did not perform well on early segments but performed well on later segments of the test.

2.5 Instrument Validation

Student performance results were used to determine reliability of the instruments: (a) across the three segments of the instrument, (b) across and within the three grades, and (c) across administrations. The instruments were validated as a measure of student achievement for teacher decision-making relative to problem solving. As stated earlier, the instruments were intended to: (a) measure growth fall to spring for a single student progressing through a school year, (b) differentiate among student performances within a single grade, and (c) differentiate

achievement across grades 6 through 8. The instruments were evaluated for their alignment to the construct of problem solving, their representation of student performance relative to this construct, and their independence from other constructs.

Approximately one-half of the teachers involved undertook an effort to teach a particular structured problem solving technique while the balance of the teaching staff made no declared effort at teaching any single technique. The problem solving materials and techniques described by the teachers were not considered by the researchers in validating the instruments, as the intent was not to validate the instructional approach but rather the measures.

3. Results

Student performance data were analyzed to determine if any student performed poorly on either segment one or segment two and performed well on a subsequent segment. The items were recoded from scaled performance scores ranging from 1-5 to dichotomous scores. Scores of 0, 1, or 2 on the scaled rubric equated to a low score for the dichotomous scale; a score of 3, 4, or 5 on the scaled rubric equated to a high score on the dichotomous scale.

Similarly, it was surmised that, without the scaffolding provided, students who did not face task one could not be expected to do well on task two while those missing task two would have difficulty with task three. For example, students who did not have a data analysis and display (task two) could not hope to correctly interpret those data (task 3). It was theorized that this difficulty was overcome using the scaffolding in each segment. Because tasks were delivered across three days, a number of students fell into the pattern of missing early segments of the test. Student performance in this instance was recoded similarly to that described above.

The data in both cases, where students performed poorly on early segments or were missing early segments altogether, indicate that there are a number of students who performed well on tasks following either poor performance on earlier tasks or no exposure to the earlier task at all. Given the nature of the tasks involved and their interdependency, this seems unlikely in the absence of the scaffolding provided. From the data returned, a total of 288 task segments across all students, forms, and administrations presented high scores after a student had received a low score (LHL=40, LLH= 119, LHH=47, and HLH=82 where L represents a low score (0 – 2) and H a high score (3 – 5)). Similarly, among students who missed one or more segments of the test, 57 segments were scored high after a missing segment (MHM=2, MMH=13, MHH=15, and HMH=27 where M represents a missing score and H a high score). A total of 345 scores showed markedly improved performance from one segment to another from 2747 total segments administered across all form and all administrations. That is, in nearly 13% of the cases students performed poorly or not at all on early segments and performed well on later segments.

3.1 Validating the Instruments

To evaluate the validity of the instruments, the research team addressed several aspects of validation including face validity, construct validity, content validity, and criterion validity. Researchers began by asking the teachers involved in the project to review the instruments for face validity (do the measures appear to be appropriate to the construct of problem-solving and for the grade level to be assessed?). It was agreed that the materials appropriately addressed the instructional goals and educational attainment of the students involved. Some concern was expressed regarding the significant amount of reading demanded of students by the assessments. It was decided that any difficulties associated with reading skill would be exposed by student performance on the instruments when compared to students' outcomes on an independent measure of reading (the state reading assessment administered in the spring of the 5^{th} grade year for the 6^{th} graders and in the spring of the 8^{th} grade year for the 8^{th} graders).

The first step, previous to addressing issues of validity, is to establish indicators of reliability. Reliability was measured across administration by comparing performances of students on counterbalanced administrations. Across all grades and across administrations, the relationship between alternate forms of the instruments was moderate, positive, and statistically significant (science form A-form B, $r(55) = .35$, $p < .05$; science form B-form A, $r(43) = .58$, $p < .05$; social studies form A-form B, $r(35) = .45$, $p < .05$; social studies form B-form A, $r(39) = .35$, $p < .05$).

Criterion validity helps to establish relationships between the constructs measured with one test and those measured with another. Unfortunately, there was no state test available in problem solving so no such correlation was possible. In the absence of such a measure, it was decided to test for divergent criterion validity relative to student scores on statewide measures of reading, writing, mathematics, and science. By correlating performance on the instruments under review with scores on the statewide tests we attempted to distinguish the construct of these problem solving instruments from the statewide tests in reading, writing, mathematics, and science.

None of the instruments at either 6th or 8th grades correlated significantly with the state writing examination. The remaining correlations, reported in Table 5, indicate that the instruments correlated with reading state examinations at both grade levels and with mathematics and science scores among 8th graders.

Table 5 *Pearson correlations of instruments with statewide testing*

	Grade 6 (spring of 5th grade year)			Grade 8 (spring administration)		
Problem Solving Measures	Reading	Science	Math	Reading	Science	Math
Fall Science	0.26*	Not Tested	0.29	0.57*	0.46*	0.36*
Fall Social Studies	0.25*	Not Tested	0.14	0.53*	0.60*	0.51*

As indication of construct validity, that is, that the four instruments were measuring the same construct, correlations were calculated across instruments within and across administrations. Because of cell size limitations, it is inappropriate to calculate correlations by grade. It was, however, possible to calculate correlations across disciplines within a single administration. Correlating science to social studies results showed relationships between science and social studies instruments by form were moderate to strong, positive, and statistically significant with the exception of science form B with social studies form B (science form A—social studies form A, $r(20) = .72$, $p < .05$; science form A—social studies form B, $r(54) = .46$, $p < .05$; science form B—social studies form A, $r(30) = .38$, $p < .05$; science form B—social studies form B, $r(13) = -.14$, $p < .05$). Note the small *n* for the final correlation.

The sensitivity of the instruments across grades and across time within grades is important to any evaluation of validity in that it affects the decisions that can be made from the results of testing using these instruments. A repeated-measures analysis of variance revealed that the student performance improved over time in both science and social studies regardless of order of administration (see Table 6). This sensitivity is indicated by the graphs in Figures 1 and 2. Note that each shows some growth within a single year by grade with the exception of science measures at the 8th grade. Also note that 7th grade students outperformed students in the 8th grade on some of the administrations.

Table 6 *Repeated Measure Analysis of Variance of Problem Solving Measures in Science and Social Studies*

	Time 1		Time 2						
	M	*SD*	*M*	*SD*	*n*	Factors	*df*	*F*	*p*
Science 1-2	6.15	1.96	6.87	2.31	75	2	1	4.84	.03
Science 2-1	6.44	2.18	7.42	1.96	55	2	1	11.26	.00
Social St. 1-2	6.44	2.19	7.59	1.90	55	2	1	19.34	.00
Social St. 2-1	7.47	2.17	8.35	2.13	43	2	1	11.19	.00

All tests are significant with p < .05.

Figure 1. Social studies problem-solving improvement by grade fall to spring administration.

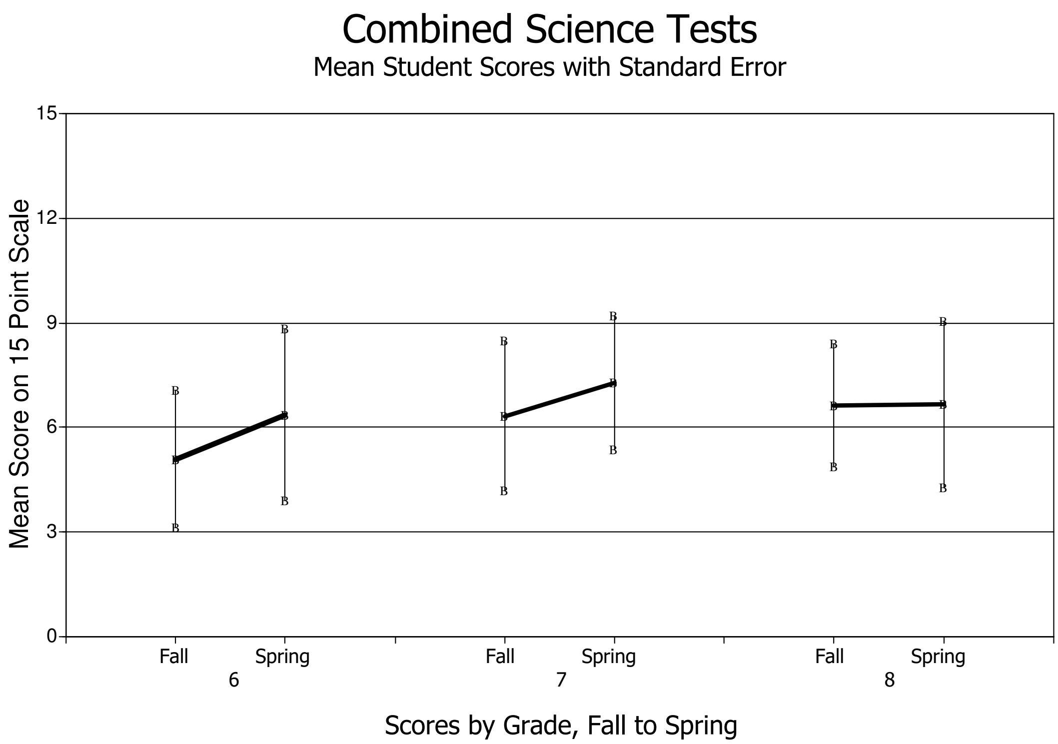
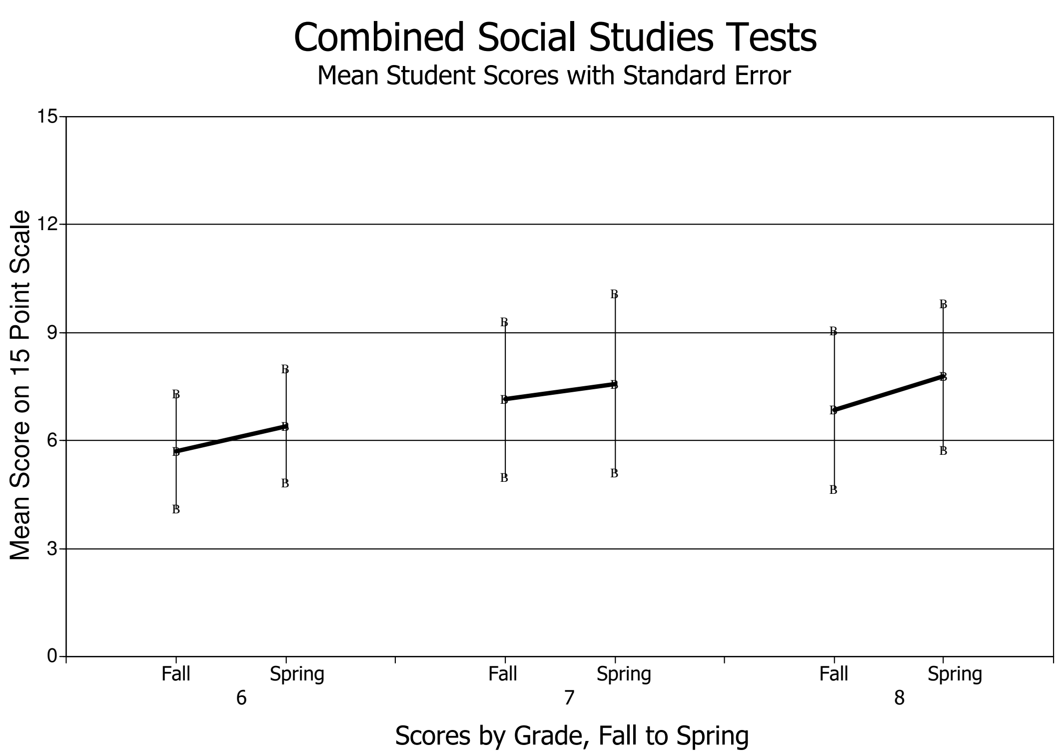


Figure 2. Social studies problem-solving improvement by grade fall to spring administration.



4. Discussion

This study represents an effort to validate the instruments used as measures of student problem-solving skills. The recoded scores show that students who missed or did not succeed on early segments of the assessments were able to succeed on later segments. Addressing face validity, teachers indicated that the instruments were appropriate to the grade level and educational attainment of students. Tests of reliability indicated that the instruments were reliable across forms and disciplines while correlations to statewide tests illustrate that, while there is no relationship between the scores on the instruments under review and statewide measures of writing skill, there is a relationship between scores on these instruments and statewide scores in reading at both 6th and 8th grades and with science and math at the 8th grade. Correlations across instruments found relationships with the exception of a single pair with only 13 test-takers in common. The instruments were found to indicate growth among students in the 6th and 7th grades while indicating flat performance among 8th graders.

The results presented above indicate that the instruments under investigation are a sound measure of problem-solving and offer scores from which valid decisions regarding student problem-solving skills can be made, thereby addressing the concerns of those suggesting that problem solving was not useful as an assessment tool (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Segers, Van den Bossche, & Teunissen, 2003) because of the "lack of rapid, valid, and reliable quantified-scoring techniques" of problem solving (Anderson, Sensibaugh, Osgood, & Mitchell, 2011, p. 1).

The data regarding student performance improvement across segments within a single measure show that the scaffolding aspect of the measures shows potential. As suggested in the scaffolding literature (Flick, 2003; Wood, Bruner, & Ross, 1979) without the scaffolding provided, students who missed or who received low scores on early segments would not have been likely to succeed on later segments. The interdependency of the segments of the instruments dictate that, without the scaffolding, students who performed poorly on early segments or who did not face early segments of the test would have faced later segments without the necessary information and would have struggled to complete the later segments at all much less with a high score.

The measures used in this investigation conformed to the four step process of problem solving described in various literature (Brophy, 1998; National Research Council, 1996, 2000) Concerns about the reading and mathematics load presented by the instruments were verified by the statistical analysis. A correlation exists between student scores on this measure and measures of student reading skill. This relationship is not surprising as the task was designed to draw on a student's skill in comprehending and interpreting a large body of text describing a problem. Additionally, the data indicate a correlation between mathematics achievement and student scores on this measure. This, too, was an issue of design. These measures were intended to indicate, to an extent, students' skill in applying mathematics concept knowledge. Student skill in both reading and mathematics were judged to be germane to the construct of problem-solving as addressed here. That is, one cannot solve this type of problem without first reading and understanding the issues involved and second addressing certain basic statistical and probabilistic issues within mathematics.

Similarly, correlations among the instruments to 8th grade statewide science testing were a positive indicator that the instruments under investigation were measuring an aspect of science reflected in the statewide science testing. The tasks also correlated across disciplines within the instruments themselves. This may indicate that problem solving is a skill independent of discipline but such a claim would require further investigation. While correlations were not strong enough to indicate that the instruments might be interchangeable across science and social studies, this may be the result of the small sample resulting from attrition. The correlation of each measure with the statewide writing examination was weak and indicated that the students' skill in writing was not a significant factor in communicating their problem solution. This was a desirable result as there was no intent that students' writing skill should be reflected in the outcome.

The data reveal that the measures are sensitive enough to measure growth in problem-solving fall to spring in both science and social studies regardless of order of administration. Variance in student performance across grades may be the result of a lack of attention to problem-solving as an instructional or curricular issue in many of the classrooms or perhaps by grade (note the relatively low and flat performance of 8th grade students).

4.1 Limitations and Future Research

While these instruments provide a basis for decision-making relative to student problem-solving skill, there are certain limitations in the content of the measures. It appears that the distinction between the two cities used in one of the social studies instruments was not adequate to provide students with a clear decision. Students, residents of one of the two cities, were often inclined to overlook social studies issues and respond with points such as, "All of my friends are here." rather than focusing on social studies or geography issues.

The sample size became an issue because of mobility among students over the course of the school year. Attrition of student participants caused limited availability of data in some conditions.

Teachers were present during the administration of the tests. This may have led to a diminution of some of the scores as teachers suggested that, not only would the scores not count for grading, but further that this was being done merely as, "a favor to the folks from the university." Some students indicated in their responses that they were unconcerned about the outcome of the testing.

It would be appropriate to extend this research beyond the middle school in an attempt to verify its sensitivity. Students in 9th or 10th grades may well perform better than did the middle school students included here. Such improvement in performance would show the sensitivity to growth important to decision-making.

As problem solving grows in prominence, developing and testing measures for effectiveness will be important in supporting the rationale for these efforts. Evidence of differential skill among learners will enhance the argument that the skill offers future value. Similarly, assessments that can be used formatively to support refinement of instruction will support instructional implementation. Additional research addressing the component aspects of problem solving, measures targeting each aspect, and scaffolds that support students as they move from aspect to aspect will strengthen the educational usefulness of the approach.

5. Conclusion

These instruments present students with the opportunity to demonstrate problem-solving skill across disciplines and to show improvement over time. Students with skill in only 1 or 2 aspects of problem-solving as defined here are supported in their efforts to accurately demonstrate their skill independent of the other aspects of problem-solving. The data reported here indicate that teacher decisions made based on the results of testing with these instruments may be valid in planning for curriculum and instructional interventions in support of problem-solving instruction.

References

Anderson, W.L., Sensibaugh, C.A. Osgood, M.P. & Mitchell, S.M. (2011). "What Really Matters: Assessing Individual Problem-Solving Performance in the Context of Biological Sciences." *International Journal for the Scholarship of Teaching and Learning,* **5**(1), 1-20.

Brophy, D.R. (1998). "Understanding, Measuring, and Enhancing Individual Creative Problem-Solving Efforts." *Creativity Research Journal*, **11**(2), 123-150.

Cho, Y. H., Caleon, I. S., & Kapur, M. (2015). Authentic Problem Solving and Learning for Twenty-First Century Learners. In Authentic Problem Solving and Learning in the 21st Century (pp. 3-16). Springer Singapore.

De León, L. (2012). "Model of Models: Preservice Teachers in a Vygotskian Scaffold." *The Educational Forum*, **76**, 144–157.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). "Effects of Problem-Based Learning: a Meta-Analysis." *Learning and Instruction,* **13**(5), 533–568.

Flick, L.B. (2003). "Teaching Science as Inquiry by Scaffolding Student Thinking." *Science Scope*, **26**(8), 34-38.

Jonassen, D. H. (2014). Assessing Problem Solving. In Handbook of Research on Educational Communications and Technology (pp. 269-288). Springer New York.

Jonassen, D. H. (2011). "Supporting Problem Solving in PBL." *The Interdisciplinary Journal of Problem-Based Learning*, **5**(2), 95-119.

Joseph, L.M. (2002). “Best Practices in Planning Interventions for Students with Reading Problems.” In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology IV*. pp 803-816. National Association of School Psychologists: Bethesda, MD.

National Research Council (2000). “How People Learn: Brain, Mind, Experience, and School.” Washington, DC: National Academy Press.

National Research Council (1996). “National Science Education Standards.” Washington, DC: National Academy Press.

Sandoval, W.A., & Reiser, B.J. (2004). “Explanation-Driven Inquiry: Integrating Conceptual and Epistemic Scaffolds for Scientific Inquiry.” *Science Education* **88**, 345-372.

Segers, M., Van den Bossche, P., & Teunissen, E. (2003). “Evaluating the Effects of Redesigning a Problem-Based Learning Environment.” *Studies in Educational Evaluation,* **29**, 315-334.

White, B.Y., & Frederiksen, J.R. (1998). “Inquiry, Modeling, and Metacognition: Making Science Accessible to all Students.” *Cognition and Instruction*, **16**(1), 3-118.

Wolf, S.E., Brush, T., & Saye, J. (2003). “Using an Information Problem-Solving Model as a Metacognitive Scaffold for Multimedia-Supported Information-Based Problems.” *Journal of Research on Technology in Education*, **35**(3), 321-341.

Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Child Psychology, Psychiatry,* **17**, 89-100.